



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A simulation of small group discussion

Citation for published version:

Padilha, E & Carletta, J 2002, 'A simulation of small group discussion', Paper presented at 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG 2002), Edinburgh, United Kingdom, 4/09/02 - 6/09/02.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A simulation of small group discussion

Emiliano G. Padilha* and Jean Carletta

ICCS – University of Edinburgh
2 Buccleuch Place, EH8 9LW, Edinburgh, UK
[emilianp,jeanc]@cogsci.ed.ac.uk

Abstract

We describe a simulation of turn-taking in small group discussion which includes behaviours from a range of modalities such as speech, gaze, facial expression, gesture, and posture. The simulation is intended to show how these behaviours contribute to the turn-taking process. The agents that participate in the discussion make probabilistic decisions about whether or not to exhibit behaviours such as speaking, making a backchannel, or shifting posture, within a general framework suggested by the existing largely descriptive literature on turn-taking. The group behaviours that characterize turn-taking models, such as turns, simultaneous speech, and competition for the floor, emerge from the individual behaviours of the agents. At this stage in the project the basic model has been designed and prototyped.

1 Introduction

Simulation is a well-established method for investigating computational models of dialogue. There have been a number of two-party simulations (Power, 1979; Houghton and Isard, 1987; Carletta, 1992; Walker, 1994; Guinn, 1996) focused on intentional structure aimed at task-oriented dialogue. Multi-party conversation or *group discussion*, on the other hand, has seldom been modelled, although much is understood about it. The best known simulation (Stasser and Taylor, 1991) reproduces only the order and distribution of turns by choosing the next speaker via two parameters: their relative talkativeness and how many turns have passed since they last spoke.

In group discussion, non-verbal behaviours

play an important part. Besides complementing the discourse and providing emotional content, they help coordinate the turn-taking. It is not surprising therefore that they are starting to be applied in conversational agents that interact with human users, as in e.g. Cassell et al. (1994), Beskow et al. (1997), Rickel and Johnson (2000). This improves the communicative efficiency and naturalness of such agents.

This paper describes a simulation of small group discussion, i.e. three to seven equal-status participants engaged in unstructured conversation. Communication in such groups proceeds as interactive dialogue, whereas bigger ones have a character more like serial monologue (Fay et al., 2000).

In this simulation, simplified behaviours on a range of modalities are reproduced: speech, gaze, head movement, facial expression, gesture, and posture. The contents and language processing of the conversation are abstracted away to focus on how these behaviours contribute to the turn-taking process. Within a framework suggested by the existing literature on turn-taking, participants in the discussion are modelled as independent agents with a set of parameterizable attributes that define likelihoods of performing the various behaviours. Group phenomena that characterize turn-taking models, such as normal and overlapped transitions, simultaneous speech, and the competition for the floor, emerge from individual behaviours, such as talk in turn, feedback, and turn-claiming gestures.

At this stage in the project the basic model has been designed, prototyped, and is being tested. We intend to evaluate it against a corpus of group discussions.

* This research was supported by a studentship from CNPq, Brazil's National Research Council.

2 Background

In group discussion, participants take *turns* at talk. During a turn, the speaker monitors whether others want to speak. This could be because he wishes to end the turn or in case someone else has a question or comment to interpose. The speaker's discourse has natural points for others to begin their turns. These points are called *transition relevance places*, or TRPs. According to Sacks et al. (1974), when the speaker selects one addressee, for instance by asking a question of him specifically, then that addressee will take the next turn. When the speaker does not select a particular addressee, leaving a *free* TRP, anyone can self-select to speak. If this does not occur, the speaker may (but need not) continue to talk.

2.1 Speaker transitions

Although these rules capture the basic properties of turn-taking, they are somewhat simplified. More than one participant can self-select to talk at a TRP. The earliest one generally takes the turn, but in simultaneous starts, the loudest usually wins (Meltzer et al., 1971), but not always. For instance, negative feedback can take precedence (Sacks et al., 1974). More specifically, the decision to continue in multiple talk has more to do with eagerness to make a contribution and the involvement in the discussion, that affect the priority given to one's own turn over the others (Oreström, 1983): i.e. some sort of confidence in oneself.

There are also *interruptions*, when someone starts to talk in the middle of the speaker's turn, between TRPs. They can happen because the interrupter has misjudged the location of a TRP, because he has anticipated what the speaker will say and wants to cut him short (Oreström, 1983), or when collaborating in an apparent "free-for-all" joint building of an idea (Edelsky, 1981). Theoretically, it could also be that the interrupter is not really listening anymore. The frequency of such interruptions depends on the size of the group (Fay et al., 2000) and in general three outcomes can be expected: either the interrupter stops in a *false-start*, or the speaker is *cut-off*, or he finishes his turn in an *overlapped transition*, talking simultaneously with the new speaker who takes the subsequent turn.

However, most speaker transitions do occur at TRPs, with a slight gap or overlap, or none at all (Sacks et al., 1974). *Gaps* are the silent intervals between transitions. They are an emergent phenomena of the group, and different from a speaker's *pause* within his turn. Mean intervals between turns in a corpus of goal-oriented dialogue were observed to be around 450–650ms (Bull and Aylett, 1998). In cases of slight gap or overlap, hearers still perceive the transition as being smooth.

Such tight timings are possible because listeners can anticipate, or *project* the TRP, chiefly from the prosodic, syntactic and semantic characteristics of the speech (Schaffer, 1983; Oreström, 1983), and also by various non-verbal behaviours. Speakers tend to look away from their interlocutors in the early stages of planning an utterance, gazing back when done planning what to say so they can monitor uptake (Kendon, 1967). They typically break mutual gaze with their interlocutors shortly after taking the turn, but sometimes maintain it a little longer before gazing away (Novick et al., 1996). Although speakers make gestures in most of their clauses (McNeill, 1992), varying with the individual, situation, and culture, they always stop them before finishing to speak (Duncan, 1972). Changes of posture occur often when speakers initiate a turn and at boundaries of discourse segments (Schefflen, 1972), which generally coincide with TRPs (Cassell et al., 2001). Listeners wanting to talk use these cues together with the speech in their attempts to take the floor.

2.2 Listener activity

While the speaker is talking, listeners provide information to the speaker about how the communication is going. *Feedback* signals at a TRP indicate that the listener does not want to speak. Positive feedback such as nods or backchannel continuers like "uh-huh," "mhm" or "right" may mean continued attention, agreement and possibly various emotional reactions, showing that the speaker can continue. Negative feedback such as a puzzled facial expression, "eh?" or "what?", on the other hand, indicates some problem in the hearing or understanding of the message. Usually, negative feedback induces the speaker to reformulate or further explain his meaning. Feedback in gen-

eral is extremely common in two-party dialogue, but is less frequent in groups (Boden, 1994) and certainly varies across cultures, gender and individuals.

Other than feedback, listeners may also indicate that they wish to speak at the next TRP. These are *turn-claiming* signals (Duncan, 1972): for instance, posture shifts accompanied or not by gestures and feedback between TRPs¹ (Beatrice, 1985), or simultaneous talk and false-starts (Oreström, 1983). All of these behaviours would tend to draw the attention of the other participants. Meanwhile, listeners maintain long gazes at the speaker interspersed by short glances away (Argyle and Cook, 1976). Speaker and listeners thus interact actively in the discussion producing a complex pattern of verbal and non-verbal behaviours.

3 The simulation

Since we are not generating actual language, with its timing, prosody, syntax and semantics that allow one to anticipate and identify TRPs, a number of simplifications have to be made. TRPs are currently provided explicitly by the speakers, as well as being announced ahead of time by a “pre-TRP” cue. This is so that listeners can decide whether they want to talk and if so, start behaving as if they are going to take the turn. Also, in order to simulate the various short intervals between turns and to distinguish precedence in multiple starts (when generally the first continues), starting turns are timestamped in fractions of a second before or after the TRP, i.e. negative or positive offsets representing overlaps and gaps, respectively.

The behaviours of the simulation fall on the following modalities:

- speech: **start** a turn (with a timestamp), **talk** in turn, **announce** a TRP (the pre-TRP cue), **arrive** at a TRP possibly **selecting** next-speaker, or make a **positive** or **negative feedback**.
- head/face: **nods** are a positive feedback, and a **puzzled expression** is a negative feedback.
- gesture: participants can **gesture**. The speaker can gesticulate throughout his

turn, while listeners can indicate they wish to speak by gesturing before a TRP.

- posture: participants can **shift posture**. Listeners can shift posture when wanting to talk, or as they start a turn. As speaker, he can shift posture again when finishing his turn, a strong indicator that he has indeed finished.
- gaze (visual monitoring): participants can **look at one another** or **at no one**. Currently, listeners all look at the speaker with occasional glances away, and he looks at the previous speaker and then looks away, gazing back when finished planning what to say.
- listening (audio monitoring): participants **pay attention to the speaker** and the speaker to the others except when busy planning what to say, in which case he pays attention **to no one** (Butterworth, 1980).

Compatible feedback from the verbal and non-verbal modalities can occur simultaneously at TRPs². Participants start to speak at a TRP, but occasionally can interrupt in the middle of a turn too. While simultaneous talk persists, as in multiple starts or at these interruptions, some or all of them can decide to stop. If everyone stops, then no one is speaking and everyone can again decide to start to speak.

3.1 Participants

Participants in the discussion are modelled as independent, autonomous agents defined by a set of constants that govern their behaviours probabilistically. Currently, the attributes we defined for each agent are:

talkativeness likelihood of wanting to talk;

transparency likelihood of producing explicit positive and negative feedbacks, and turn-claiming signals;

confidence likelihood of interrupting, and continuing to speak during simultaneous talk;

interactivity the mean length of turn segments between TRPs;

¹The same signals used for feedback can also be used as *turn-precursors*: e.g. nodding before starting a turn.

²Clearly, non-verbal feedback alone is only effective when its producer holds speaker’s gaze, but people give it even when they do not: e.g. in telephone conversation.

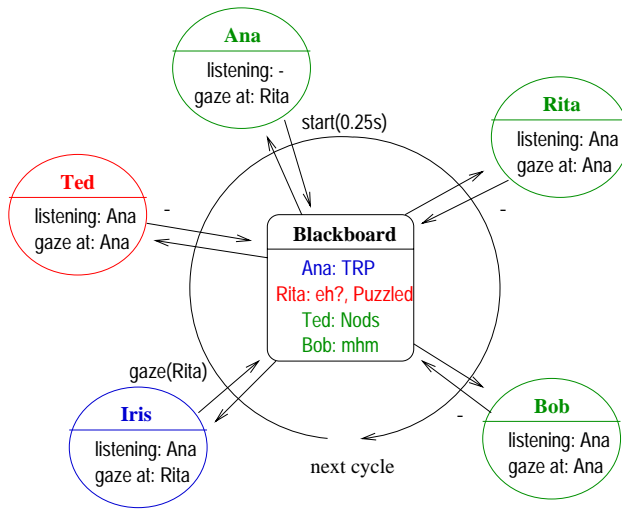


Figure 1: One cycle of conversation.

verbosity likelihood of continuing the turn after a TRP at which no one self-selected.

These values can be generated automatically for a group of agents at the start of the simulation (for example, based on a given mean and standard deviation), or set individually. They are a means of obtaining a variable pattern of behaviour for the group.

3.2 Architecture

In order to simulate independent agents performing sometimes simultaneous behaviours, the discussion is run in a loop with a clock *cycle* of an arbitrary length³ during which agents read and write behaviours to a blackboard (figure 1). This represents the environment of the conversation: everything that is said and done. At each cycle, agents read the blackboard to *perceive* what has happened in the previous one. All behaviours in a cycle are meant to be “simultaneous.”

With this framework, speaking turns and gesturing spread across several cycles. Group phenomena such as gaps and simultaneous talk at speaker transitions, too, can stretch for more than one cycle. Other, more instantaneous behaviours such as feedback, shifts of gaze and posture occur within only one cycle in the simulation.

³A good choice for its length is 500ms, or half-a-second. Thus turn starts, which vary in this range from the TRP (positively, or negatively in overlaps), can be contained in one such cycle.

3.3 Processing

At each cycle, an agent reads the behaviours from the blackboard, decides what to do and writes back the new behaviours. Decisions are all probabilistic, based on the likelihoods given by his attributes. The agent has some variables recording the discourse context, so that they know, for instance, when they are in the middle of a turn. His decisions in a cycle are as follows:

- If (no one is speaking)
 - test **talkativeness** to start to speak here
 - if so, start with a random interval
 - test **transparency** to shift posture.
- If (listening to a single speaker)
 - look at him; occasionally, look away
 - if (read the pre-TRP)
 - test **talkativeness** to decide to start
 - if so, test **transparency**
 - to make turn-claiming signals now
 - mark next cycle as the TRP
 - if (at a TRP *and* decided to start)
 - or (at a TRP *and* was selected)
 - start with a random interval
 - test **transparency** to shift posture
 - if (at a *free* TRP *and* not going to start)
 - test **transparency** to do feedback
 - if (anywhere else)
 - test **talkativeness and confidence** to start to speak, i.e. to interrupt.
- If (started simultaneously at a TRP)
 - test **confidence and** who started first to decide whether to continue.
- If (speaking simultaneously, *and* not planning)
 - test **confidence** whether to continue.
- If (speaking alone in a turn)
 - use **interactivity** to set the segment length
 - decide when to gesture and gaze-away
 - decide when planning stage ends
 - gaze back at interlocutor at that point
 - if (at the last cycle before the TRP)
 - decide whether to select next-speaker
 - if (arrived at a TRP *and* no one started)
 - test **verbosity** to continue talking.

Some explanation is in order. Turn-claiming signals are gesture, posture shift, nods and/or “mhm.” They are decided individually according to the agent’s transparency. In multiple starts, the agents test confidence reduced appropriately by their precedence in time with regard to the others. The only other combined

test uses talkativeness and confidence in deciding whether to interrupt.

Regarding the speaker, the segment length from TRP to TRP is based on his interactivity: the higher it is the shorter the segments. Other minor decisions could be naturally based on his verbosity: frequency of selecting next-speaker and the length of the planning stage. Still others, such as gestures and gaze-away, are fixed.

As for feedback, the higher the transparency attribute the more likely the agent is to make explicit signals, and in more than one modality at once. The simulation chooses negative cases in a fixed proportion. Feedback is only performed at a *free* TRP, i.e. when the speaker did not select-next⁴. This selection is indicated at the last cycle before the TRP. Only the selected agent, reading this, will decide to start at the TRP.

3.4 Example

Figure 2 presents a log of three turns in the simulation. For the sake of space, only three participants are shown, each with three columns:

- speech: with turn starts shown only by the timestamp from the TRP, and the pre-TRP cue indicated by pTRP;
- “body” behaviours: GESTures, POSTure shifts, NODS and PUZZled facial expressions;
- gaze: at other participants or at no one.

In this example, Iris begins to talk shifting posture and gesticulating in the middle. She passes a TRP being acknowledged by the others. At the next TRP, Rita slightly overlaps her (by 50ms). It is not possible to know whether Iris was intending to continue after the TRP, but since someone started, she stops. Ana also starts to speak at that TRP but gives in, especially since she began later than Iris. Iris continues with talk in turn, shortly ending at a TRP. Rita gives positive feedback at the TRP but Ana takes the turn perceivably later (more than one cycle away). At Ana’s first TRP, Rita complains, causing her to start a new turn to address the problem. The timestamp of that

⁴Negative feedback certainly can occur when someone is selected, but this would mean the speaker has to make another select-next turn with the same addressee. We are leaving these contextual complexities for later.

Iris (body,gaze)			Ana (body,gaze)			Rita (body,gaze)		
		Rita			Rita			Ana
+ .25s	POST	Rita			Rita			Ana
talk		-			Iris			Iris
talk	GEST	-			Iris			Iris
talk	GEST	-			Rita			Iris
pTRP	GEST	Rita			Iris			Iris
talk		Rita			Iris			Iris
TRP		Rita		NODS	Iris	mhm	NODS	Iris
talk		-			Iris			Iris
talk		-			Iris			Iris
pTRP		-			Iris			Iris
talk		-			Iris			Iris
TRP		Rita	+ .30s		Iris	- .05s	GEST	Iris
		Rita			Rita	talk	GEST	Iris
		Rita			Rita	talk		-
		Rita			Rita	pTRP		-
		Rita			Rita	talk		Iris
mhm	NODS	Rita			Rita	TRP		Iris
		Ana	+ .25s	POST	-			Iris
		Ana	talk		-			Ana
		Ana	pTRP		Rita			Ana
		Ana	talk		Rita			Ana
		Ana	TRP		Rita	eh?	PUZZ	Ana
		Ana	+ .10s		Rita			Ana
		:		:	:			:

Figure 2: Excerpt of three turns of conversation.

start is in this case relative to the TRP in her own talk. Again it is not possible to know whether Ana would have continued or stopped at that TRP had no negative feedback been produced. And so the simulation continues until it is stopped by the user or reaches an indicated number of turns.

4 Evaluation

The most usual type of formal evaluation requires a corpus against which properties of the simulation can be measured. Although there is some group corpus data available, it is still extremely limited. Corpora like that are extremely expensive to produce. Moreover, this is a new research area, and research concerning any behavioural phenomenon must begin with descriptive analysis, with quantitative corpus work following later. Our simulation synthesizes a substantial body of largely descriptive analysis pertaining to turn-taking in small groups, showing how the different behaviours work together. It is thus a preliminary step compared to work currently being undertaken in more mature areas.

We have access to two sources of corpus data. The first, the corpus used by (Fay et al., 2000), consists of group discussions among undergraduates. There are ten groups each of five and ten participants, each of which used a circumscribed scenario which limited the topics in-

volved and allowed them to determine the relationship among group size, discussion structure, and influence. Discussions were recorded on audiotape using two non-directional microphones. The second consists of two group discussions, one of size five and the other of size eight, using Fay’s scenario but recorded to give better information for all modalities. These groups were audio-recorded on individual, synchronized tracks. The participants wore baseball hats marked with contrasting arrows, making their head direction and gestures apparent from a ceiling-mounted video recording.

Evaluation of a model such as the one underlying our simulation is performed by showing that the proposed model fits the real data better than some simpler, baseline model. This is the approach used by (Stasser and Taylor, 1991). In this case, the proposed model might, for instance, predispose the group to dyadic turn patterns by giving a starting advantage to the agent at whom the speaker gazes when approaching a TRP and making it likely for the speaker to gaze at the last speaker. Whether or not this elaboration is necessary would be tested by checking whether a model containing it can be made to fit the data better than one that excludes it. The properties of small group discussion which we feel it would be most useful for our model to explain are:

- the distribution of turns among speakers;
- dyadic patterning in turn sequences;
- turn and turn segment length;
- rate, location, and length of simultaneous speech and interruptions;
- placement and frequency of positive and negative feedback in the various modalities;
- placement and frequency of the explicit signals that a listener wishes to take a turn.

Clearly the amount of data that we have with clear recording of all modalities is limited enough to make evaluation difficult for some properties of the model. However, at this early stage, even measurement and synthesis of a coherent, first-pass model is a useful research aim. Our results could be used simply to inform upon the rate and rough placement of the behaviours

we study. For many animators wishing to create naturalistic agents, this may be enough information to improve their results.

5 Conclusions

We described a simulation of small group discussion that attempts to reproduce patterns of turn-taking. Actual language contents and processing that drive a dialogue were ignored in order to focus on how turn-taking behaviours in a range of modalities constrain and shape the conversation, that is, those contents. This does not mean that the turn-taking drives the contents, but that the contents do not directly affect turn-taking behaviour.

This simulation improves upon previous work in a number of aspects. Simulations of group discussion such as (Stasser and Taylor, 1991) traditionally run in a central loop where all decisions are made; participants do not take them individually. In this work, they are independent agents capable of behaving simultaneously. They are modelled after a set of attributes that define complex individual ‘personalities.’ Turn-taking patterns thus emerge from the independent decisions they make. They perform a range of multi-modal behaviours that result in talk in turn, feedback and turn-claiming signals. They engage in various types of speaker transitions including interruption, overlap and selection-of-next. And finally they analyse the turn in progress for when to speak or to perform feedback, and whether others are doing this too, hence determining who takes the turn at the TRP.

The behaviours presented here are *placeholders* or abstractions for future elaboration of the various complex phenomena that surface in multi-party conversation. So they generalize a range of subtly different cases. In real conversations, an interjection like “oh!” or “ah!”, for example, can have different meaning and function (as backchannel, turn-precursor, etc) depending on their context, intonation and facial expression accompanying it. Gestures may vary in type, scope and energy and even postures communicate moods and situations: e.g. slouching, stiff, twisted (Blatner, 2002). All these subtleties show how complex non-verbal communication is and how it contributes to the conversation.

We intend to proceed with some elaborations to improve the simulation. One first possibility is to replace explicit TRP indicators by pauses and changes in intonation, volume and rate of speech, thus making agents analyse the actual cues for turn-taking. This might be further enhanced by generation of simple sentence patterns for the contents of the conversation, containing topics, speech-act types and obligations. On top of this, various types of performance phenomena, like pre-starts, tag questions, hesitations, and self-interruptions, might be relevant once a model of the speaker is employed, such as the one proposed by Levelt (1989). All of this can further affect decisions on whether to speak and listen, whether to interrupt and continue simultaneous talk, to perform feedback, where to gaze, and so on.

This work is basic research. It is intended to inform about multi-party conversation, in particular small group discussion, by emulating multi-modal simultaneous behaviours of multiple independent agents in such setting. It could also inform practical work on, for instance, embodied conversational agents, more natural “barge-in” for spoken dialogue systems and to help coordination of video-conferences. One way in which it can help is by emulating the ways by which people anticipate when new utterances are expected (when someone is going to talk) by their behaviours while listening. Another is by generating appropriate behaviours to coordinate speaker transitions and keep the conversation flowing. Thus we expect to provide a first step in the development of better models of agents that can naturally interact in multi-party conversation.

References

- M. Argyle and M. Cook. 1976. *Gaze and mutual gaze*. Cambridge University Press, Cambridge, MA.
- G. Beattie. 1985. The threads of discourse and the web of interpersonal involvement. *Bulletin of the British Psychological Society*, 38:169–175.
- Jonas Beskow, Kjell Elenius, and Scott McGlashan. 1997. OLGA - A dialogue system with an animated talking agent. In *Proceedings of the Eurospeech'97*, pages 1651–1654, Rhodes, Greece.
- A. Blatner. 2002. *About nonverbal communications*. Available at <http://www.blatner.com/adam/level2/nverb1.htm>.
- D. Boden. 1994. *The Business of Talk: Organizations in Action*. Blackwell, Oxford, England.
- M. Bull and M. Aylett. 1998. An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. In *Proceedings of the IC-SLP'98*, volume 4, pages 1175–1178, Sydney, Australia.
- B. Butterworth. 1980. Some constraints on models of language production. In *Speech and talk*, volume 1 of *B. Butterworth (ed.) Language production*, London. Academic Press.
- J. Carletta. 1992. *Risk-taking and Recovery in Task-oriented Dialogue*. Ph.D. thesis, University of Edinburgh, Edinburgh, Scotland.
- J. Cassell, N. Badler C. Pelachaud, B. Achorn M. Steedman, B. Douville T. Becket, and M. Stone S. Prevost. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of the SIGGraph'94 Computer Graphics Annual Conference*, volume 1, pages 413–420, New Orleans.
- J. Cassell, C. Sidner Y. Nakano, T. W. Bickmore, and C. Rich. 2001. Non-verbal cues for discourse structure. In *Proceedings of the 41st Annual Meeting of the ACL*, pages 106–115, Toulouse, France.
- S. Duncan. 1972. Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, 23(2):283–293.
- C. Edelsky. 1981. Who's got the floor? *Language and Society*, 10(3):383–421.
- N. Fay, S. Garrod, and J. Carletta. 2000. Group discussion as interactive dialogue or serial monologue: the influence of group size. *Psychological Science*, 11(6):487–492.
- C. I. Guinn. 1996. Mechanisms for mixed-initiative human-computer collaborative discourse. In A. Joshi and M. Palmer, editors, *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics*, pages 278–285, San Francisco, USA.
- G. Houghton and S. Isard. 1987. Why to speak, what to say and how to say it: modelling lan-

- guage production in discourse. In *P. Morris (ed.) Modelling Cognition*, pages 249–267.
- A. Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.
- W. J. M. Levelt. 1989. *Speaking: from Intention to Articulation*. MIT Press, Boston.
- D. McNeill. 1992. *Hand and Mind: What gestures reveal about thought*. The University of Chicago Press, Chicago.
- L. Meltzer, W. Morris, and D. Hayes. 1971. Interruption outcomes and vocal amplitude: Explorations in social psycho-physics. *Journal of Personality and Social Psychology*, 18:392–402.
- D. G. Novick, B. Hansen, and K. Ward. 1996. Coordinating turn-taking with gaze. In *Proceedings of the ICSLP'96*, volume 3, pages 1888–1891, Philadelphia.
- B. Oreström. 1983. *Turn-taking in English Conversation*. Lund Studies in English. LiberFörlag Lund, Malmö, Sweden.
- R. Power. 1979. The organization of purposeful dialogues. *Linguistics*, 17:107–152.
- J. Rickel and W. Johnson. 2000. Task-oriented collaboration with embodied agents in virtual worlds. In *Cassell, J., J. Sullivan, S. Prevost et al. (eds.) Virtual Worlds, in Embodied Conversational Agents*, Boston. MIT Press.
- H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language*, 50(4):696–735.
- D. Schaffer. 1983. The role of intonation as a cue to turn-taking in conversation. *Journal of Phonetics*, 11:243–259.
- A. E. Schefflen. 1972. *Body language and the social order*. Prentice Hall, Englewood Cliffs, NJ.
- G. Stasser and L. A. Taylor. 1991. Speaking turns in face-to-face discussions. *Journal of Personality and Social Psychology*, 60(5):675–684.
- M. A. Walker. 1994. Discourse and deliberation: testing a collaborative strategy. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 1205–1211, Kyoto, Japan.